(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 978 844 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
09.02.2000 Bulletin 2000/06

(51) Int. Cl.[7]: G11C 11/56

(21) Application number: 98830491.1

(22) Date of filing: 07.08.1998

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(71) Applicant:
STMicroelectronics S.r.l.
20041 Agrate Brianza (Milano) (IT)

(72) Inventors:
- Calligaro, Cristiano
  27020 Torre d'Isola (PV) (IT)
- Rolandi, Paolo
  27058 Voghera (PV) (IT)
- Gastaldi, Roberto
  20041 Agrate Brianza (MI) (IT)
- Torelli, Guido
  27016 S. Alessio con Vialone (PV) (IT)

(74) Representative: Mittler, Enrico
c/o Mittler & C. s.r.l.,
Viale Lombardia, 20
20131 Milano (IT)

(54) Sensing arrangement for a multilevel semiconductor memory device

(57) A multilevel memory device comprises an array of multilevel memory cells (M_{1j} - M_{kj}, M_{1z} - M_{kz}) arranged in rows (WL_1 - WL_k) and columns (BL_j, BL_z), each memory cell being capable of being programmed in $m = 2^n$ ($n > 1$) distinct programming states, and a sensing arrangement for sensing the memory cells, the sensing arrangement comprising at least (m - 1) reference columns ($BL_{ref,i}$, $BL_{ref,h}$) of memory cells. The reference columns comprises a number of memory cells substantially identical to the number of memory cells of each column of the array, a smaller number of memory cells ($M_{ref,i}$, $M_{ref,h}$) of each reference column being multilevel reference memory cells programmed in a respective reference programming state and activatable for sinking a respective reference current ($I_{R,0}$, $I_{R,1}$, $I_{R,2}$), the remaining larger number of memory cells of each reference column being dummy non-conductive memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$) structurally identical to the reference memory cells and to the memory cells of the array.

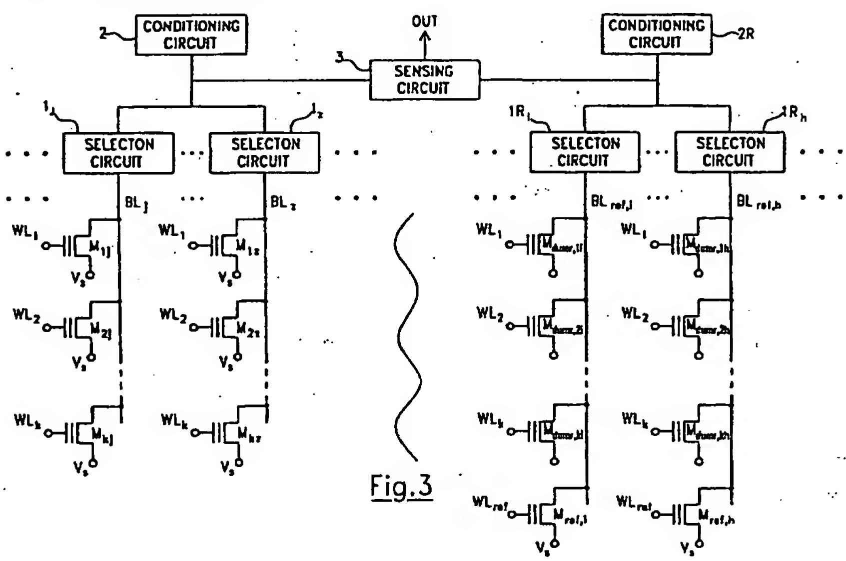


Fig.3

EP 0 978 844 A1

Description

[0001] The present invention relates to multilevel semiconductor memory devices, particularly multilevel non-volatile semiconductor memory devices. More particularly, but not exclusively, the invention relates to multilevel non-volatile semiconductor memory devices wherein the memory cells comprise MOS transistors and the programming of the memory cells is achieved by means of a change in the threshold voltage of the MOS transistor. Still more particularly, but not exclusively, the invention relates to multilevel non-volatile memory devices such as EPROMs, EEPROMs, Flash EEPROMs, wherein the memory cells comprise floating-gate MOS transistors.

[0002] In multilevel memory devices each memory cells can take one among m distinct predetermined programming states. This means, if $m = 2^n$, the memory cell can store $n = \log_2 m$ bits of information. Clearly, a multilevel memory device with a given area has a memory capacity much higher than an equivalent bilevel memory device of the same area.

[0003] Conventionally, a non-volatile memory cell of the memory cell array or matrix (in the following called for the sake of simplicity matrix memory cell) is read (or "sensed") by comparing the threshold voltage of such a memory cell with the threshold voltage of so-called reference memory cells shortly, reference cells, properly programmed. The reference memory cells are structurally identical to the matrix memory cells: this assures a good matching of characteristics between the reference and the matrix memory cells, and improves the precision of the matrix cell sensing operation. The matrix memory cells can take one of m distinct threshold voltage levels; if the threshold voltages of the reference memory cells are properly set, the comparison between the threshold voltage of the matrix memory cells and that of the reference memory cells univocally provides the information corresponding to the data stored in the matrix memory cells.

[0004] In the practice, sensing of a given matrix memory cell is performed by biasing in substantially identical conditions both the matrix memory cell and the reference memory cells. This means, the voltages applied to the drain, gate, source and bulk terminals of the matrix memory cell are substantially identical to those applied to the respective terminals of the reference memory cells. Then, the current sunk by the matrix memory cell is compared with the currents sunk by the reference memory cells. Given the biasing conditions, there is a one-to-one relationship between the threshold voltage of a memory cell, be it matrix or reference, and the current sunk by the same; clearly, a higher threshold voltage will correspond to a lower current sunk.

[0005] In the case of bilevel memory cells, the matrix cell current, in the predetermined biasing conditions, can be either zero (matrix cell programmed so to have a high threshold voltage value, programming state *0*) or substantially equal to a predetermined nominal value (matrix cell programmed so to have a low threshold voltage, programming state *1*). The reference memory cell can be programmed so as to sink, in the predetermined biasing conditions, the same current sunk by a matrix cell in the *1* programming state. It is thus simple to determine, by current comparison, the programming state of the matrix cell. And it is also simple to program the reference memory cells.

[0006] In the case of multilevel memory cells with m distinct possible values for the threshold voltage, the matrix cell current is to be compared with $m - 1$ properly set reference currents. Fig. 1 shows the case of a four-level matrix memory cell which can assume four different threshold voltage values, corresponding, in the prescribed sense biasing conditions, to four matrix cell currents $I_{C,0}$ to $I_{C,3}$. For each matrix cell current there is also indicated the corresponding value of a two-bit digital code stored in the matrix memory cell. $I_{R,0}$ to $I_{R,2}$ are the three values of the reference cell currents to be compared with the matrix cell current. Advantageously, the reference current values fall in the middle of the intervals between adjacent matrix cell currents, i.e., ideally, $I_{R,i} = 1/2\,(I_{C,i} + I_{C,i+1})$, with $i = [0; m-2]$. Whichever the programming state of the matrix memory cell, comparison of the matrix cell current with the reference cell currents provides univocally the information stored in the matrix memory cell. By way of example, if the data stored in the matrix memory cell is "10", the matrix cell current I_{cell} is equal to $I_{C,2}$. The comparison will give a negative result for every reference cell current I_{ref} higher than $I_{R,1}$, while it will give a positive result for every reference cell current I_{ref} lower than or equal to $I_{R,1}$. A similar situation occurs for every other possible value of the memory cell current I_{cell}, and for every value of m different from 4.

[0007] Setting the reference cell currents in the middle between adjacent matrix cell currents provides the best margin against data retention degradation and operating condition changes.

[0008] It is to be noted that even if $m - 1$ distinct reference current values suffice for a correct sensing of the data stored in a multiple level memory cell with m distinct programming levels, nothing prevents from using an additional reference current value, either higher than $I_{C,m-1}$ or lower than $I_{C,0}$ (clearly in the case $I_{C,0} > 0$), or even both of said additional reference values, even if this is not strictly necessary.

[0009] Figure 2 shows schematically a sensing arrangement for reading multilevel memory cells in a memory matrix. This arrangement is clearly a straightforward extension of the conventional sensing arrangement for bilevel memory cells. In the drawing, BL_j is representative of one of a plurality of matrix bit lines. M_{1j} to M_{kj} are multilevel matrix memory cells of the memory matrix belonging to the j-th bit line BL_j. Matrix memory cells M_{1j} to M_{kj} have their drain electrodes connected to bit line BL_j. WL_1 to WL_k are the word lines of the mem-

ory matrix, and each memory cell of bit line BL_j has the control gate electrode connected to a respective one of word lines WL_1 to WL_k. When a word line is selected, its voltage is brought to a prescribed positive value, suitable for optimizing the sense operation; the remaining word lines are normally kept at ground. The source electrodes of the matrix memory cells are connected to a low-impedance source line Vs, substantially a voltage generator; conventionally, in EPROM devices source line Vs is the ground; in Flash EEPROMs Vs is a line common to all the memory cells of a memory sector, and it can be switched between a read voltage (normally, ground) and an erase voltage.

[0010] A number of matrix bit lines are connected to a selection circuit 1 which selects one among all the said number of matrix bit lines connected thereto according to address signals not shown. Selection circuit 1 allows to connect the selected matrix bit line to a conditioning circuit 2 and to a sensing circuit 3. Conventionally, the conditioning circuit 2 includes circuits for biasing the selected bit line, i.e. for biasing the drain electrodes of the matrix memory cells connected to the selected bit line, as well as preset circuits such as for example precharge circuits. This is well known to the skilled reader, and will not be explained in further detail.

[0011] At the reference side, *m* - 1 reference bit lines $BL_{ref,1}$ to $BL_{ref,m-1}$ of reference memory cells (e.g., $M_{ref,1i}$ to $M_{ref,ki}$ for reference bit line $BL_{ref,i}$) are each one connected to a respective selection circuit $1R_1$ to $1R_{m-1}$. Selection circuits $1R_1$ to $1R_{m-1}$ allow to connect one of the reference bit lines to a conditioning circuit 2R and to the sensing circuit 3. The conditioning circuit 2R performs similar functions as conditioning circuit 2.

[0012] Sensing circuit 3 performs the comparison between the matrix cell current flowing through the selected matrix bit line, and the reference cell current flowing through the selected reference bit line. Output terminal OUT of the sensing circuit 3 provides the information corresponding to the data stored in the selected matrix memory cell sensed.

[0013] Each reference memory cell of a reference bit line has the drain electrode connected to the reference bit line, the control gate electrode connected to a respective one of the word lines WL_1 - WL_k, and the source electrode connected to the source line Vs. This assures a good matching between the matrix memory cells to be sensed and the reference memory cells that are used for comparison, since when a matrix memory cell of a given word line is to be sensed, reference memory cells belonging to the same word line are used. In order to further improve the matching, the reference bit lines can be located inside the matrix of memory cells.

[0014] All the reference memory cells that, such as reference memory cells $M_{ref,1}$ to $M_{ref,k}$, belong to a same reference bit line such as $BL_{ref,i}$, are programmed at the same level. This means, considering for example the case of four-level memory cells shown in Fig. 1, that three reference bit lines at least are to be provided, a first reference bit line made of reference cells all programmed at the level corresponding to current $I_{R,0}$ a second reference bit line made up of reference cells all programmed at the level corresponding to current $I_{R,1}$, and a third reference bit line made up of reference cells all programmed at the level corresponding to current $I_{R,2}$.

[0015] This technique optimize the matching between the reference memory cells and the matrix memory cells. The reference memory cells are in fact structurally identical to the matrix memory cells, are fabricated simultaneously with and by means of the same process steps as the matrix memory cells and therefore are affected by same deviations from ideality. Additionally, during the sensing phase, one of the two input terminals of the sensing circuit (comparator) is connected to the matrix bit line to which the matrix memory cell to be sensed belongs, and the other input terminal is connected to one of the reference bit lines, not to a single reference cell. This assure a pretty good dynamic matching, since the capacitive loads seen by the sensing circuit at the matrix side are quite the same as those seen at the reference side. The whole sensing operation, inclusive of the pre-charge phase and the very sensing phase, is therefore optimized. The same beneficial effects are encountered in the "program verify" step in electrically programmable memory devices, step that provides for sensing a programmed matrix cell under programming to detect if it has really been programmed at the desired level.

[0016] However, this solution poses some problems when applied to multilevel memory cells. Actually, a large number of reference memory cells, specifically all the reference cells of a given reference bit line, are to be programmed at the same programming level. The problems involved are, from one hand, the lenghty programming step of all the reference memory cells and, from the other hand, the difficulty in precisely programming all the reference memory cells of a given reference bit line to the very same programming level. The time to be dedicated to the programming of the reference memory cells increases with the precision of their programming, and with the increase of the number of programming levels of the matrix cells.

[0017] Clearly, the use of a single reference memory cell in lieu of a whole reference bit line of reference memory cells would solve the above-mentioned problems, but at the expense of a good matching between the circuit branch at the memory cell matrix side and the circuit branch at the reference memory cell side.

[0018] In view of the state of the art described, it has been an object of the present invention that of providing a multilevel memory device suitable for assuring a good matching during sensing between the matrix bit line containing the memory cell to be sensed and the reference bit lines, but at the same time overcoming the problems of the conventional ones.

[0019] According to the present invention, such an

object is attained by means of a multilevel memory device, comprising an array of multilevel memory cells arranged in rows and columns, each memory cell being capable of being programmed in $m = 2^n$ ($n > 1$) distinct programming states, and a sensing arrangement for sensing the memory cells, the sensing arrangement comprising at least (m - 1) reference columns of memory cells, characterized in that each one of said reference columns comprises a number of memory cells substantially identical to the number of memory cells of each column of the array, a smaller number of memory cells of each reference column being multilevel reference memory cells programmed in a respective reference programming state and activatable for sinking a respective reference current, the remaining larger number of memory cells of each reference column being dummy non-conductive memory cells structurally identical to the reference memory cells and to the memory cells of the array.

[0020] The features and advantages of the present invention will be made apparent by the following detailed description of some particular embodiments thereof, illustrated by way of non-limiting examples only in the annexed drawings, wherein:

Figure 1 diagramatically shows a conventional, optimum placement of reference levels for the sensing of four-level memory cells;
Figure 2 shows a sensing arrangement for a multilevel memory device which is a straightforward extension of a conventional sensing arrangement for bilevel memory cells;
Figure 3 shows a sensing arrangement for a multilevel memory device according to a first embodiment of the present invention;
Figure 4 shows a sensing arrangement for a multilevel memory device according to a second embodiment of the present invention;
Figure 5 shows a sensing arrangement for a multilevel memory device according to a third embodiment of the present invention;
Figure 6 shows in top-plan view the layout of a portion of a memory cell array of the multilevel memory device shown in Figure 4;
Figure 7 shows a schematic cross-section along line VII-VII of Figure 6;
Figure 8 shows the same portion of memory cell array shown in Figure 6, but obtained by a different manufacturing process;
Figure 9 shows a sensing arrangement for a multilevel memory device according to a fourth embodiment of the present invention;
Figure 10 shows a sensing arrangement for a multilevel memory device according to a fifth embodiment of the present invention; and
Figure 11 shows a sensing arrangement for a multilevel memory device according to a sixth embodiment of the present invention.

[0021] Referring to Fig. 3, a sensing arrangement for a multilevel memory device according to a first embodiment of the present invention is shown. In the drawing, two matrix bit lines BL_j, BL_z of multilevel memory cells M_{1j} to M_{kj} and M_{1z} to M_{kz} are shown, representative of a plurality of matrix bit lines of a memory matrix or memory sector of the multilevel memory device. Each matrix bit line is connected to a respective selection circuit such as 1_j, 1_z, operating the selection of the respective bit line. The outputs of the selection circuits 1_j, 1_z are connected to a conditioning circuit 2 and to a sensing circuit 3. Each matrix bit line contains a number of matrix multilevel memory cells equal to the number of word lines WL_1 - WL_k of the memory matrix or memory sector.

[0022] At the reference side, two reference bit lines $BL_{ref,i}$ and $BL_{ref,h}$ representative of a plurality of m-1 (at least) reference bit lines are shown. Each reference bit line is connected to a respective selection circuit $1R_i$, $1R_h$ operating the selection of the respective reference bit line; selection circuits $1R_i$, $1R_h$ are similar to those provided for the matrix bit lines. The outputs of the selection circuits $1R_i$, $1R_h$ are supplied to a conditioning circuit 2R and to the sensing circuit 3.

[0023] Each reference bit line contains a number of dummy memory cells $M_{dumr,1i}$ - $M_{dumr,ki}$ and $M_{dumr,1h}$ to $M_{dumr,kh}$ equal to the number of word lines WL_1 to WL_k of the memory matrix or sector. Additionally, each reference bit line contains a further reference multilevel memory cell $M_{ref,i}$ and $M_{ref,h}$. The dummy memory cells have a drain electrode connected to the respective reference bit line, a control gate electrode connected to a respective word line WL_1, WL_k, but their source electrode is left floating. Differently, the reference memory cell $M_{ref,i}$, $M_{ref,h}$ of each reference bit line has a drain electrode connected to the respective reference bit line, a control gate electrode connected to a reference word line W_{ref} distinct from the matrix word lines WL_1 - WL_k, and a source electrode connected to the same source line Vs as the source electrodes of the matrix memory cells. The dummy memory cells are structurally identical to the matrix memory cells and the reference memory cells. For each reference bit line, the respective reference memory cell is programmed at a prescribed level so that, under the prescribed bias conditions, the current flowing through the reference bit line is that required for performing the comparison with the current flowing through the selected matrix bit line; for example, in the case of four-level memory cells, as reported in Fig. 1, three reference bit lines are at least necessary, and the respective three reference memory cells of such reference bit lines will have to be programmed to levels corresponding to the reference current values $I_{R,0}$, $I_{R,1}$ and $I_{R,2}$. Differently, the dummy cells of the reference bit lines are not to be programmed in any particular programming level, since being their source electrodes floating the dummy cells will never contribute to the current flowing through the respective reference

bit line.

[0024] During the sensing phase, the matrix bit line containing the addressed matrix memory cell is selected by the respective selection circuit, and it is connected to the sensing circuit. Also the matrix word line containing the addressed matrix memory cell is selected. Thus, the addressed memory cell will sink a current that depends on the particular programming level.

[0025] The reference word line WL_{ref} is also selected, and the reference bit lines are connected, one at a time, to the sensing circuit 3. The only current flowing through the reference bit line is that contributed by the respective reference memory cell; in fact, even if the dummy cell of the reference bit line having its control gate electrode connected to the selected matrix word line will be selected, such a dummy cell does not contribute any current being its source electrode floating.

[0026] Clearly, connection of only one reference bit line at a time to the sensing circuit 3, that is a so-called "serial sensing" of the matrix memory cells, is just one of the possible sensing approaches which have been proposed in the art for multilevel memory; other approaches are for example the parallel and the mixed serial-parallel. It is apparent that the present invention applies as well to the other sensing approaches not only to the serial sensing described for simplicity.

[0027] If desired, in order to achieve a better matching between the matrix memory cells and the reference memory cells, the reference bit lines can be placed inside the memory matrix, adjacent to matrix bit lines.

[0028] For each reference bit line, the respective reference memory cell can be placed anywhere, either at the top, or at the bottom of the reference bit line, as well as in any other position, such as at the middle of the reference bit line.

[0029] It appears that thanks to the present arrangement, each reference bit line is physically and electrically substantially identical to a generic matrix bit line. In fact, each reference bit line includes substantially the same number of memory cells as any matrix bit line: the mismatch between the matrix bit lines and the reference bit lines is 1/k, where k is the number of word lines of the memory matrix, i.e. the number of matrix memory cells attached to each matrix bit line. Thus, the capacitive loads seen by the sensing circuit at the reference side will be substantially equal to those seen at the matrix side.

[0030] However, compared to the conventional arrangement shown in Fig. 2, it is not necessary to program all of the memory cells connected to a reference bit line to a same programming level. Indeed, only the reference memory cells $M_{ref,i}$, $M_{ref,h}$ are to be programmed to the respective programming level, because these are the only cells that contribute to the reference current.

[0031] Sensing arrangements for multilevel memory devices in accordance with a second and third embodiments of the present invention are shown in Figs. 4 and 5. In the arrangement of Fig. 4, differently from the previous one, a dummy memory cell $M_{dum,j}$, $M_{dum,z}$ with floating source electrode is additionally connected to each matrix bit line BL_j, BL_z. The control gate electrodes of the dummy memory cells of the matrix bit lines are connected to the reference word line WL_{ref}. In this way, the matching between the matrix bit lines and the reference bit lines is improved and it is almost perfect, since the former includes exactly the same number of memory cells as the latter. Clearly, being the source electrode of the dummy memory cells of the matrix bit lines floating, they do not contribute any current to the matrix bit line current.. In the arrangement of Fig. 4 the reference memory cells $M_{ref,i}$, $M_{ref,h}$ attached to the reference bit lines and the dummy memory cells $M_{dum,j}$, $M_{dum,z}$ attached to the matrix bit lines are located at one end of the matrix and reference bit lines. However, the reference memory cells and the dummy memory cells could be located anywhere along the respective bit lines. For example, in the arrangement of Fig. 5 the dummy memory cells and the reference memory cells are located in the middle of the respective bit lines.

[0032] As an alternative embodiment, for optimizing the matching between the matrix bit lines and the reference bit lines it is possible, instead of adding a dummy memory cell to each matrix bit line, to eliminate one dummy memory cell from each reference bit line, so that the number of memory cells connected to the matrix bit lines and to the reference bit lines is always equal to k.

[0033] Fig. 6 is a schematic top-plan view of a portion of a memory cell array of a multilevel memory device implementing the sensing arrangement of Fig. 4. Fig. 7 is a schematic cross-sectional view showing two adjacent matrix memory cells belonging to matrix bit line BL_j and respectively connected to matrix word lines WL_a and WL_b. However, since the reference memory cells and the dummy memory cells all have identical structure, Fig. 7 can be though to refer to any pair of adjacent cells in the memory cell array or in the reference bit lines.

[0034] In Fig. 6, two matrix bit lines BL_j and BL_z and two reference bit lines $BL_{ref,i}$ and $BL_{ref,h}$ are shown, representative of all the matrix bit lines and reference bit lines, respectively. The bit lines, shown in dashed lines, are made of metal and run parallely to each other. Also shown in Fig. 6 are portions of five word lines WL_a to WL_e, and the reference word line WL_{ref}. The word lines and the reference word line are strips made of an upper polysilicon level, and run substantially parallely to each other in the orthogonal direction compared to that of the bit lines. The word lines form the control gate electrodes of the memory cells. Where the word lines cross the bit lines, under the word lines floating gate electrodes FG made of a lower polysilicon layer are provided.

[0035] Between alternate pairs of word lines (e.g., between word lines WL_a and WL_b, WL_c and WL_d, WL_e and WL_{ref}) the bit lines contact N+ regions 10 forming a

common drain region for the pairs of adjacent cells. Each cell also have an N+ source region 11 which is merged to a respective N+ source line 12 running parallely to the word lines. A plurality of N+ source lines 12 are provided, among alternate pairs of word lines (e.g., word lines WL_b and WL_c, WL_d and WL_e, etc.). Metal source lines Vs running parallely to the bit lines are interspersed among the bit lines, e.g. one metal source line Vs every *N* bit lines with *N* normally equal to 16 or 32. The metal source lines Vs contact the N+ source lines 12 to bias the latter.

[0036] As visible from Fig. 6, in order to implement the sensing arrangement of Fig. 4 all the dummy memory cells M_{dumr} (corresponding to the dummy memory cells $M_{dumr,1i}$, $M_{dumr,2i}$,...and $M_{dumr,1h}$, $M_{dumr,2h}$,... of Fig. 4) connected to the reference bit lines $BL_{ref,i}$ and $BL_{ref,j}$ have the source electrodes 11 floating; this is achieved by interrupting the N+ source lines 12, so that the source electrodes 11 of the dummy memory cells M_{dumr} are not electrically connected to the metal source lines Vs. The same is for the dummy memory cells $M_{dum,j}$ and $M_{dum,z}$ provided in the matrix bit lines BL_j and BL_z. Instead, the reference memory cells $M_{ref,i}$ and $M_{ref,h}$ of the reference bit lines $BL_{ref,i}$ and $BL_{ref,h}$ have the source electrodes 11 connected to an N+ source line 12 that is contacted by the metal source lines Vs.

[0037] Even if the layout of Fig. 6 explicitly refers to the sensing arrangement of Fig. 4, it is straightforward for the skilled technician to derive a similar layout for the other sensing arrangements shown in Figs. 3 and 5.

[0038] A known technique for the manufacturing of memory cell arrays is that called "Self-Aligned Source" (SAS), described for example in US Patent No. 5,103,274. This technique provides first of all for forming over the semiconductor substrate 100 parallel strips of N+ doped active area, which at the end of the manufacturing process will be located under the metal bit lines. The strips of N+ doped active area are separated by strips of a thick field oxide. After having defined the word lines of the memory cell array, a photolithographic step using a mask called SAS mask is applied to the memory cell array; the SAS mask defines strips (referred to as SAS regions or SAS windows) perpendicular to the strips of field oxide, corresponding to those strips that, at the end of the process, will be the N+ source lines. An etching process is then performed for removing the field oxide along such predefined strips, i.e. the SAS windows. After this photolitographic step, an N type dopant is implanted into these predefined strips to form the N+ source lines of the memory cells.

[0039] Figure 8 shows how the conventional SAS process can be easily adapted to form a memory cell array conformal to the one shown in Fig. 6, i.e. a memory cell array implementing the sensing arrangements according to the present invention.

[0040] Fig. 8 shows the same portion of memory cell array shown in Fig. 6, but obtained by a different manufacturing process . For the sake of clarity, in Fig. 8 the SAS windows have been highlighted. The conventional SAS photolithographic mask is modified to provide the results shown in Fig. 8. Instead of having continuous windows in the SAS mask extending substantially along the whole matrix parallely to the word lines, windows 30 in the SAS mask are only provided where it is desired to form the N+ source lines 12 shown in Fig. 6, i.e. for those memory cells that shall have their source electrode 11 connected to the metal source lines Vs. During the following etching step, field oxide regions 50 at both sides of the source region of the dummy memory cells are not removed. Then, an N type dopant is implanted to form the N+ source regions of the memory cells; however, as far as the dummy memory cells are concerned, their source regions will remain isolated due to the presence at both sides thereof of the field oxide regions, masking the underlying semiconductor substrate from the implanted dopant. This occurs for all the dummy cells, i.e., both for the dummy cells belonging to the matrix memory bit lines and for the dummy cells belonging to the reference bit lines.

[0041] Figs. 9, 10 and 11 shows sensing arrangement for multilevel memory devices corresponding to alternative embodiments of those depicted in Figs. 3, 4 and 5, respectively.

[0042] Specifically, in Fig. 9, that corresponds to Fig. 3, a sensing arrangement is shown wherein the dummy memory cells of the reference bit lines are replaced by memory cells $M_{dHV,1i}$ - $M_{dHV,ki}$ and $M_{dHV,1h}$ - $M_{dHV,kh}$ that instead of having their source electrodes floating, are programmed at a level corresponding to a threshold voltage sufficiently high to make such cells non-conductive at the voltage levels normally applied to the selected word lines WL_1 - WL_k.

[0043] Fig. 10, corresponding to Fig. 4, similarly shows how not only the dummy memory cells of the reference bit lines are memory cells programmed to have a high threshold voltage, as in Fig. 9, but also the dummy cells attached to the matrix bit lines in order to optimize the matching with the reference bit lines are replaced by memory cells $M_{dVH,j}$, $M_{dVH,z}$ having their source electrodes connected to the source line Vs, but however programmed at a level sufficiently high to make such cells non-conductive at the normal voltage levels applied to the word line WL_{ref} for the selection of the reference memory cells $M_{ref,i}$, $M_{ref,h}$ of the reference bit lines.

[0044] Similarly, Fig. 11 shows the same sensing arrangement of Fig. 10, but with the real reference memory cells $M_{ref,i}$, $M_{ref,h}$ and the high-threshold voltage cells of the matrix bit lines located at a different position inside the bit lines, e.g. in the middle thereof, instead at one end.

[0045] The alternative embodiments shown in Figs. 9, 10 and 11 are simpler, from the layout point of view, than their counterparts of Figs. 3, 4 and 5, respectively. In fact, for the implementation of the sensing arrangements of Figs, 3, 4 and 5 it is necessary, as explained

above, to modify the layout of the memory cell array, so as to leave the source electrodes of the dummy memory cells floating. Instead, the embodiments of Figs. 9, 10 and 11 do not require any layout modification, but only a suitable programming of those memory cells that are to be always non-conductive, so as to rise the threshold voltage thereof. This programming step is to be performed before the programming step of the reference memory cells.

[0046] All the above-described embodiments allow to simply obtain a good matching between the matrix bit lines and the reference bit lines, but without the need of accurately programming quite a large number of reference memory cells, as instead required by the sensing arrangements according to the prior art.

[0047] The area overhead introduced by the sensing arrangements according to the invention is quite small. Let's consider for example the case of a memory device with a memory matrix divided in blocks. Let's also suppose that it is desired to introduce a group of reference bit lines in every block: this situation is the most expensive in terms of device area, but nonetheless it is the one assuring the best matching between the matrix bit lines and the reference bit lines. If the reference bit lines were absent, one source metal line Vs every 16 or 32 matrix bit lines would normally be provided. The layout arrangement would be the following: one source metal line Vs, 16 or 32 matrix bit lines, one source metal line, etc. When the $(m - 1)$ reference bit lines are inserted, a good arrangement could be the following: one source metal line Vs, 16 or 32 matrix bit lines, one source metal line Vs, 16 or 32 matrix bit lines....one source metal line Vs, $(m - 1)$ reference bit lines, one source metal line Vs, 16 or 32 matrix bit lines, etc. In this way, the memory cell array is substantially not altered from the electrical viewpoint, since the distance between the source electrode of the matrix cells and the source metal line does not change. It appears that the overhead is only given by one additional source metal line Vs and $(m - 1)$ reference bit lines for each block of the memory array.

[0048] No area overhead at all is instead introduced by using the sensing arrangements of Figures 9, 10 and 11, which only require a proper programming of the dummy cells.

[0049] In the previous description, reference has been made for simplicity to a memory cell array with memory cells having source electrodes connected to a common source line Vs, However, it is clear and evident that the present invention applies as well to memory devices having the memory cell array divided in memory sectors, such as sectorized Flash EEPROMs; as known, in this case only the memory cells belonging to a same memory sector have the source electrodes connected to each other, and to a respective memory sector source line distinct from those of the other memory sectors. Thus, all the drawings and the related description can be straightforwardly extended to a sectorized memory device, simply by considering source line Vs as a memory sector source line.

Claims

1. Multilevel memory device, comprising an array of multilevel memory cells (M_{1j} - M_{kj}, M_{1z} - M_{kz}) arranged in rows (WL_1 - WL_k) and columns (BL_j, BL_z), each memory cell being capable of being programmed in $m = 2^n$ $(n > 1)$ distinct programming states, and a sensing arrangement for sensing the memory cells, the sensing arrangement comprising at least $(m - 1)$ reference columns ($BL_{ref,i}$, $BL_{ref,h}$) of memory cells, characterized in that each one of said reference columns comprises a number of memory cells substantially identical to the number of memory cells of each column of the array, a smaller number of memory cells ($M_{ref,i}$, $M_{ref,h}$) of each reference column being multilevel reference memory cells programmed in a respective reference programming state and activatable for sinking a respective reference current ($I_{R,0}$, $I_{R,1}$, $I_{R,2}$), the remaining larger number of memory cells of each reference column being dummy non-conductive memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$) structurally identical to the reference memory cells and to the memory cells of the array.

2. Multilevel memory device according to claim 1, characterized in that said smaller number of memory cells of each reference column comprises one reference memory cell ($M_{ref,i}$, $M_{ref,h}$).

3. Multilevel memory device according to claim 2, characterized in that the dummy non-conductive memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$) of the reference columns are each one connected to a respective one of the rows (WL_1 - WL_k) of the array, while the reference memory cells ($M_{ref,i}$, $M_{ref,h}$) are connected to a reference row line (WL_{ref}) activatable for activating the reference memory cells.

4. Multilevel memory device according to claim 3, characterized in that the reference memory cells ($M_{ref,i}$, $M_{ref,h}$) are located at one end of the respective reference column ($BL_{ref,i}$,$BL_{ref,h}$).

5. Multilevel memory device according to claim 3, characterized in that the reference memory cells ($M_{ref,i}$,$M_{ref,h}$) are located among the dummy memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$) of the respective reference column ($BL_{ref,i}$,$BL_{ref,h}$), particularly in the middle of the respective reference column.

6. Multilevel memory device according to any one of claims 3 to 5, characterized in that each one of said columns (BL_j,BL_z) of the array comprises, in addi-

tion to the respective multilevel memory cells (M_{1j} - M_{kj},M_{1z} - M_{kz}), an additional dummy non-conductive memory cell ($M_{dum,j}$,$M_{dum,z}$), so that the columns of the array comprise a number of memory cells exactly equal to the number of memory cells of the reference columns ($BL_{ref,i}$,$BL_{ref,h}$).

7. Multilevel memory device according to claim 6, characterized in that the dummy memory cells ($M_{dum,j}$,$M_{dum,h}$) of the columns (BL_j,BL_z) of the array are connected to said reference row line (WL_{ref}).

8. Multilevel memory device according to any one of the preceding claims, characterized in that said reference columns ($BL_{ref,i}$,$BL_{ref,h}$) are located inside the array of multilevel memory cells.

9. Multilevel memory device according to any one of the preceding claims, characterized in that the multilevel memory cells (M_{1j} - M_{kj}, M_{1z} - M_{kz}), the dummy memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$; $M_{dum,j}$,$M_{dum,h}$) and the reference memory cells ($M_{ref,i}$,$M_{ref,h}$) comprise MOS transistors having a drain electrode connected to the respective column or reference column, a control electrode connected to a respective row or reference row line, and a source electrode.

10. Multilevel memory device according to claim 9, characterized in that the source electrodes of the memory cells (M_{1j} - M_{kj}, M_{1z} - M_{kz}) and of the reference memory cells ($M_{ref,i}$,$M_{ref,h}$) are connected to a common source line (Vs).

11. Multilevel memory device according to claim 9, characterized in that said array of multilevel memory cells comprises at least two memory sectors, the source electrodes of the memory cells and of the reference memory cells of one memory sector being connected to a common sector source line distinct from a sector source line of the other memory sectors.

12. Multilevel memory device according to claim 10 or 11, characterized in that the source electrode of the dummy memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kh}$; $M_{dum,j}$,$M_{dum,h}$) are electrically floating, to make the dummy memory cells non-conductive.

13. Multilevel memory device according to claim 10 or 11, characterized in that the source electrode of the dummy memory cells ($M_{dumr,1i}$ - $M_{dumr,ki}$, $M_{dumr,1h}$ - $M_{dumr,kj}$; $M_{dum,j}$,$M_{dum,h}$) is connected to said source line (Vs), and the dummy memory cells are programmed at a programming state suitable for making the dummy cells non-conductive when their control electrode and drain electrode are biased with voltages normally used in sensing operations.

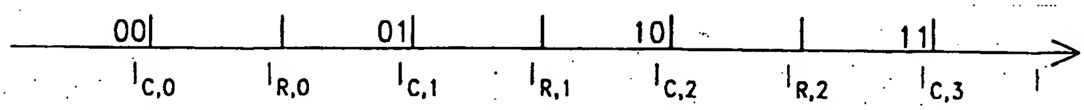


Fig. 1
(PRIOR ART)

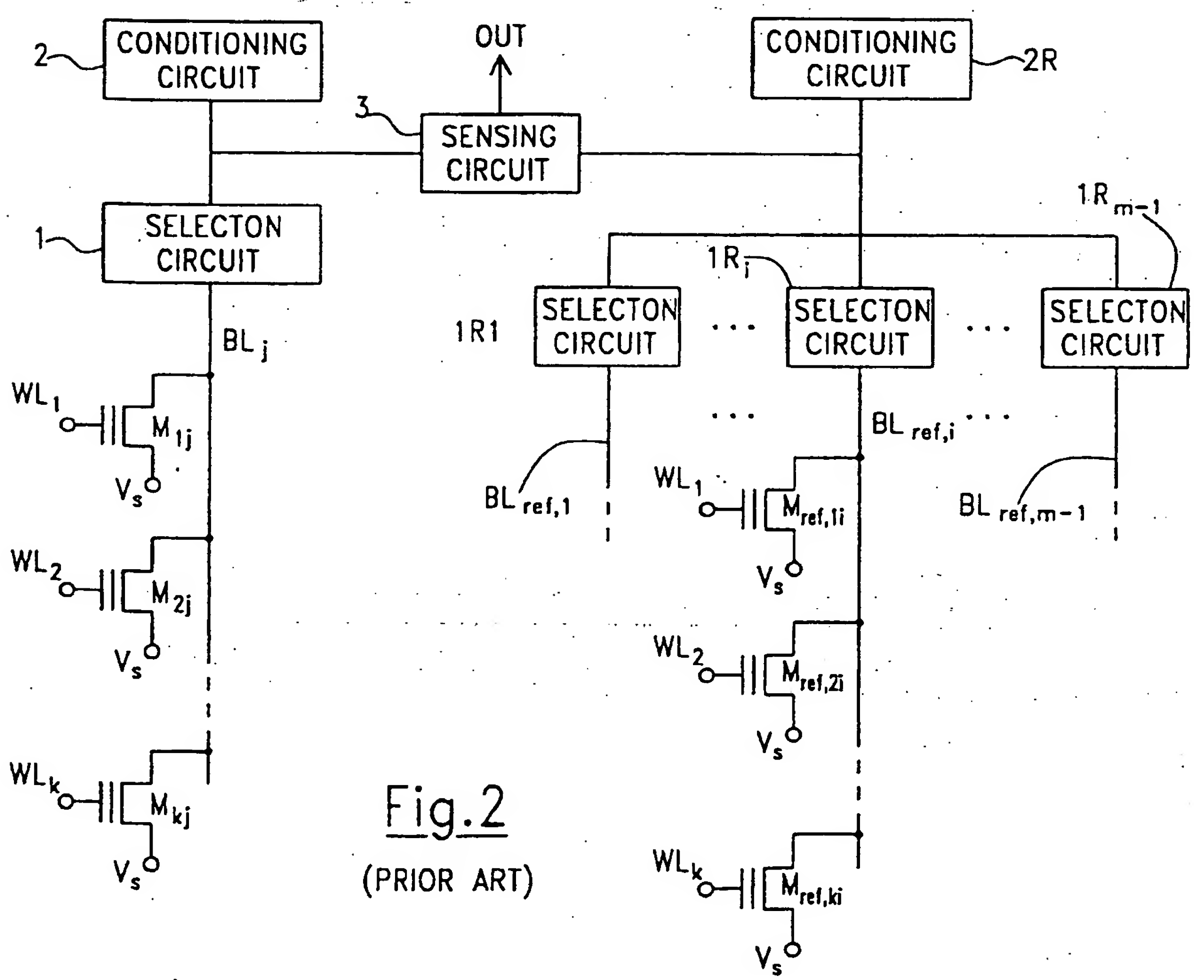


Fig. 2
(PRIOR ART)

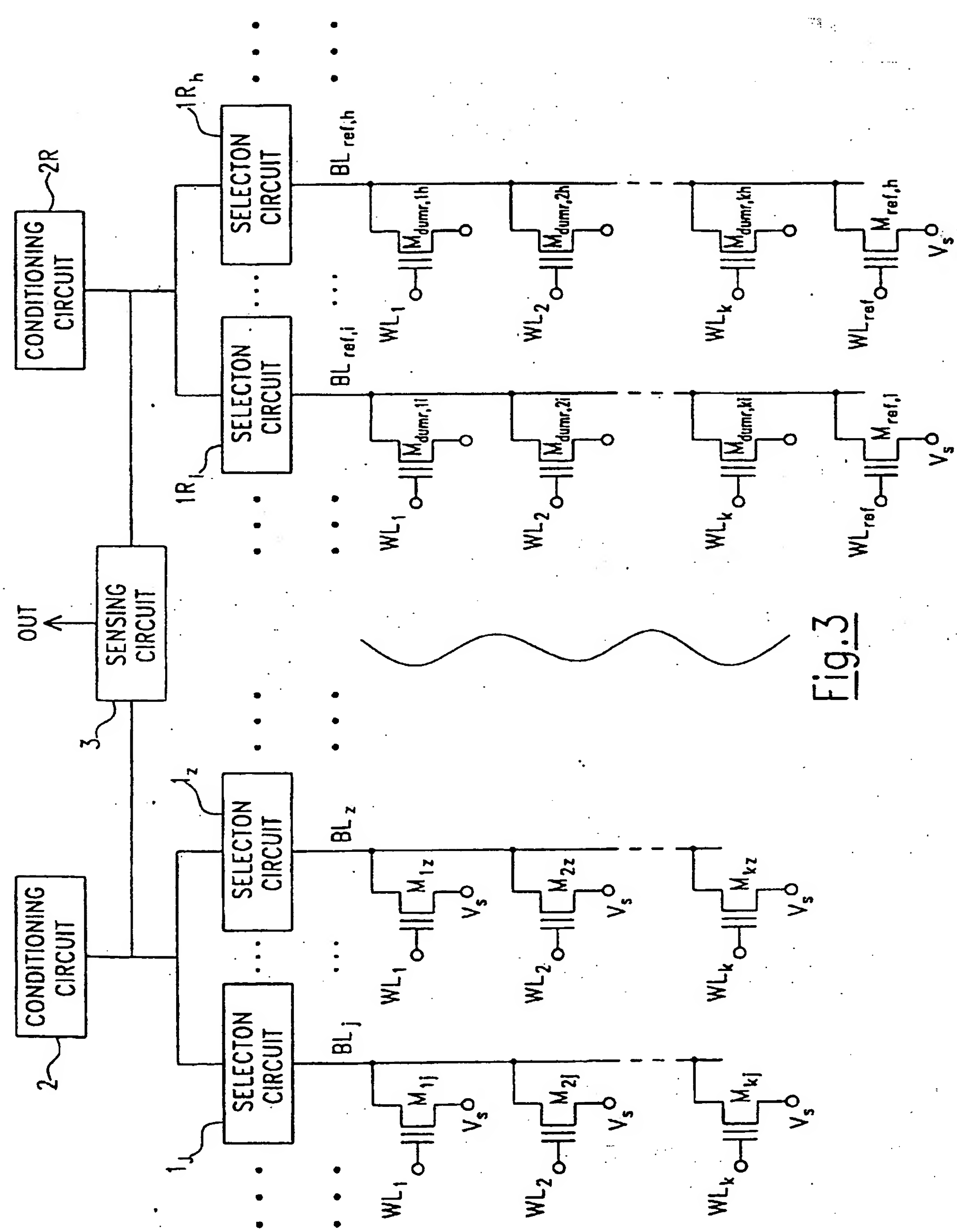
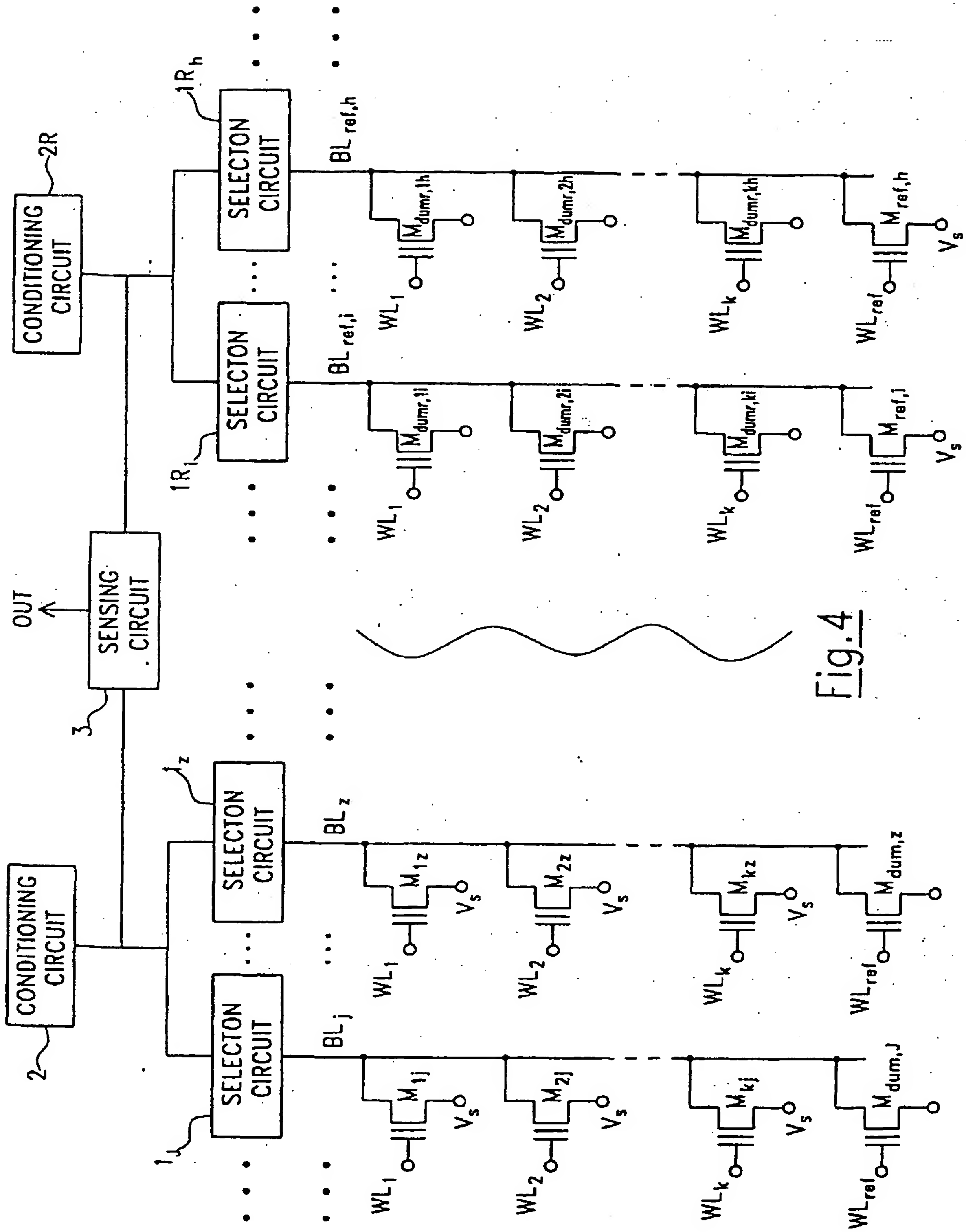

CONDITIONING CIRCUIT
2R
CONDITIONING CIRCUIT
2
SENSING CIRCUIT
3
OUT
SELECTON CIRCUIT
1Rh
1Ri
1z
1j
BLref,h
BLref,i
BLz
BLj
Mdumr,1h
Mdumr,2h
Mdumr,kh
Mref,h
Mdumr,1i
Mdumr,2i
Mdumr,ki
Mref,i
M1z
M2z
Mkz
M1j
M2j
Mkj
WL1
WL2
WLk
WLref
Vs

Fig.3



CONDITIONING CIRCUIT
2R
CONDITIONING CIRCUIT
2
SENSING CIRCUIT
3
OUT
SELECTON CIRCUIT
$1R_h$
$1R_1$
1_z
1_j
$BL_{ref,h}$
$BL_{ref,i}$
BL_z
BL_j
$M_{dumr,1h}$
$M_{dumr,2h}$
$M_{dumr,kh}$
$M_{ref,h}$
$M_{dumr,1i}$
$M_{dumr,2i}$
$M_{dumr,ki}$
$M_{ref,i}$
M_{1z}
M_{2z}
M_{kz}
$M_{dum,z}$
M_{1j}
M_{2j}
M_{kj}
$M_{dum,j}$
WL_1
WL_2
WL_k
WL_{ref}
V_s

Fig. 4

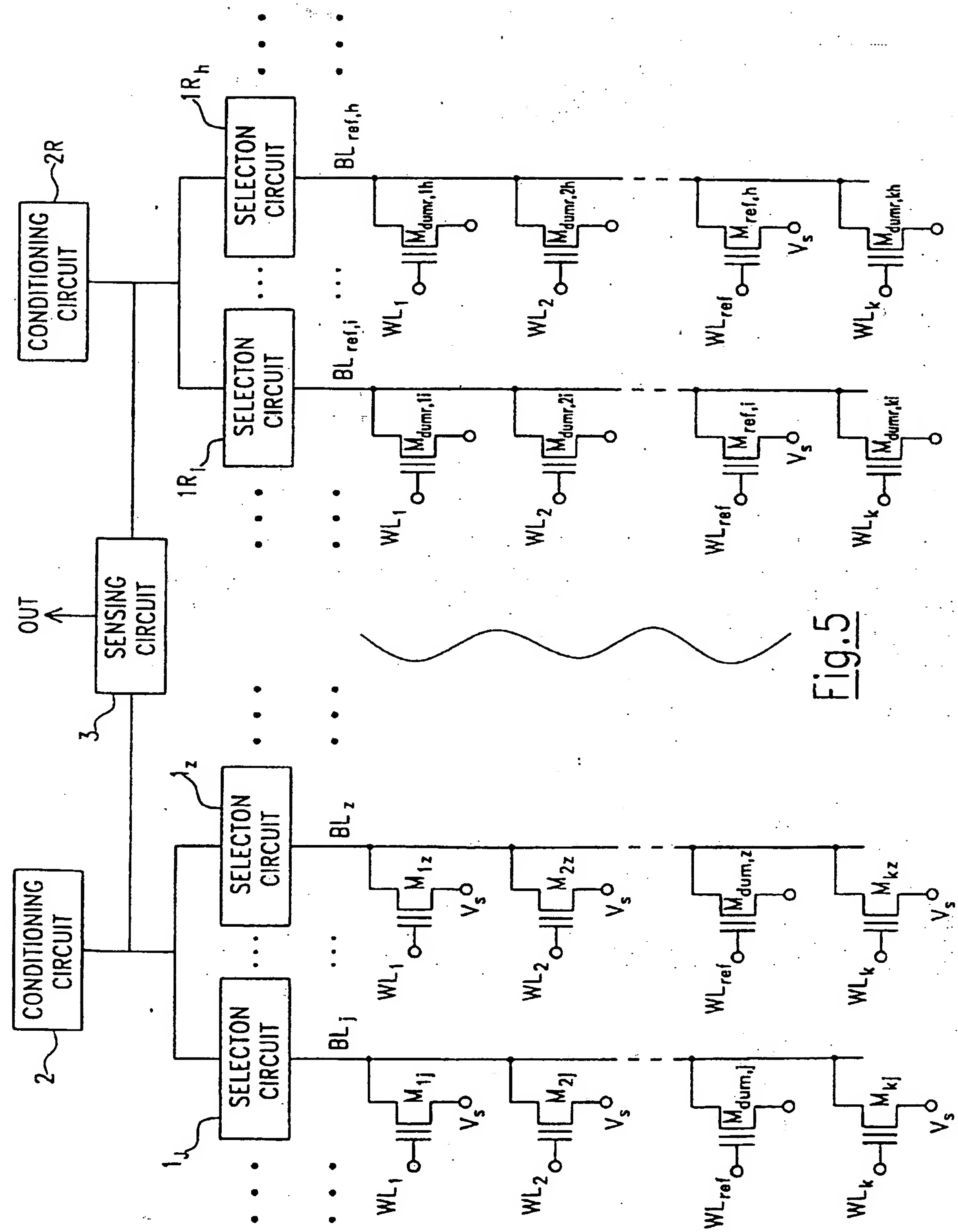

CONDITIONING CIRCUIT
2R
SELECTON CIRCUIT
1R_h
BL_ref,h
SELECTON CIRCUIT
1R_i
BL_ref,i
SENSING CIRCUIT
OUT
3
CONDITIONING CIRCUIT
2
SELECTON CIRCUIT
1_z
BL_z
SELECTON CIRCUIT
1_j
BL_j
WL_1
WL_2
WL_ref
WL_k
M_dumr,1h
M_dumr,2h
M_ref,h
M_dumr,kh
M_dumr,1i
M_dumr,2i
M_ref,i
M_dumr,ki
M_1z
M_2z
M_dum,z
M_kz
M_1j
M_2j
M_dum,j
M_kj
V_s

Fig.5

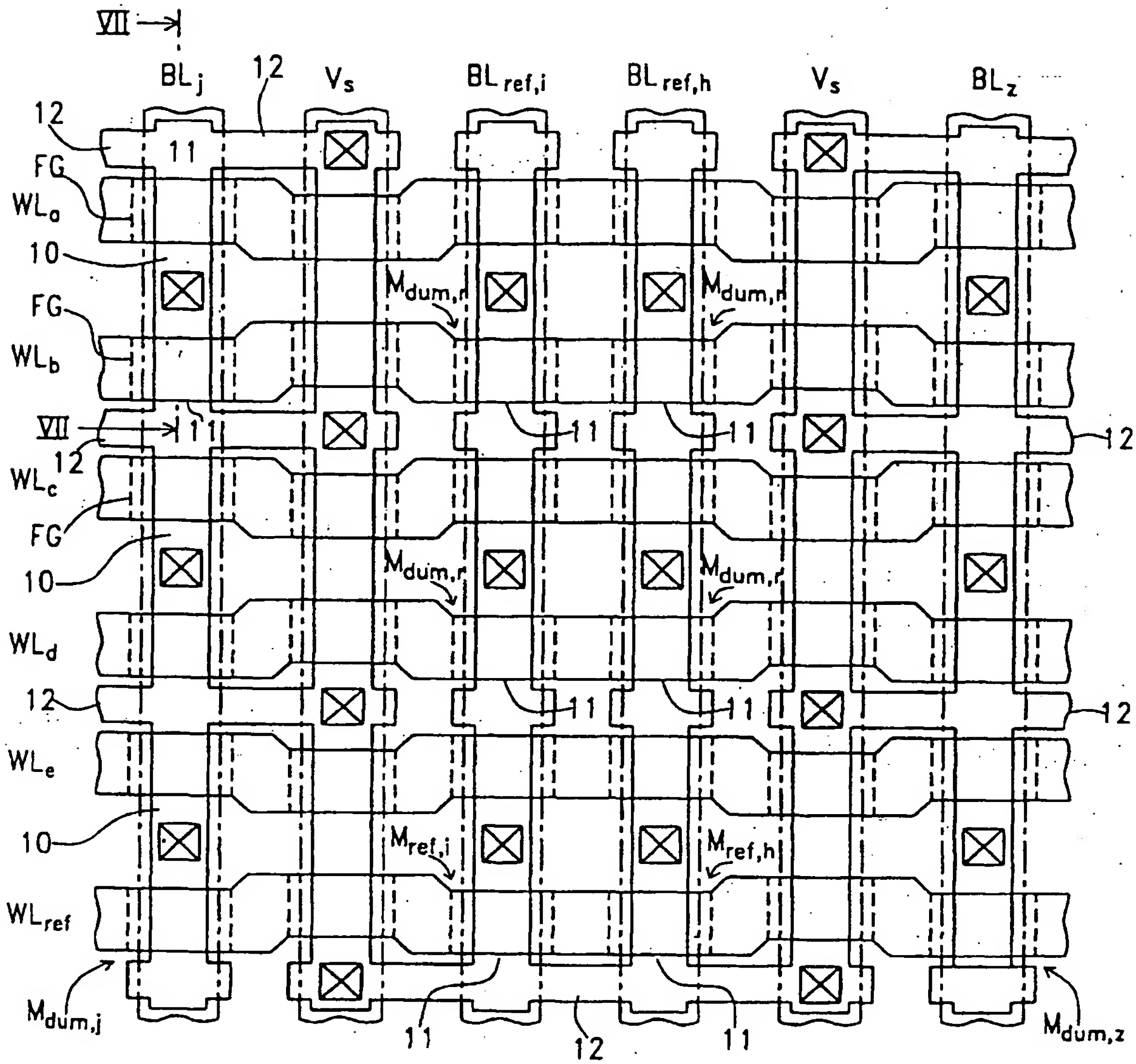

VII
BLj
12
Vs
BLref,i
BLref,h
Vs
BLz
12
11
FG
WLa
10
Mdum,r
Mdum,r
FG
WLb
VII
11
11
11
12
WLc
12
FG
10
Mdum,r
Mdum,r
WLd
12
11
11
12
WLe
10
Mref,i
Mref,h
WLref
Mdum,j
11
12
11
Mdum,z

Fig.6

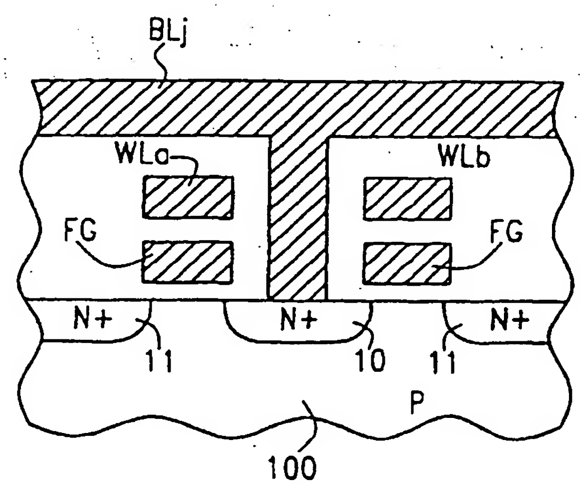

BLj
WLa
WLb
FG
FG
N+
N+
N+
11
10
11
P
100

Fig.7

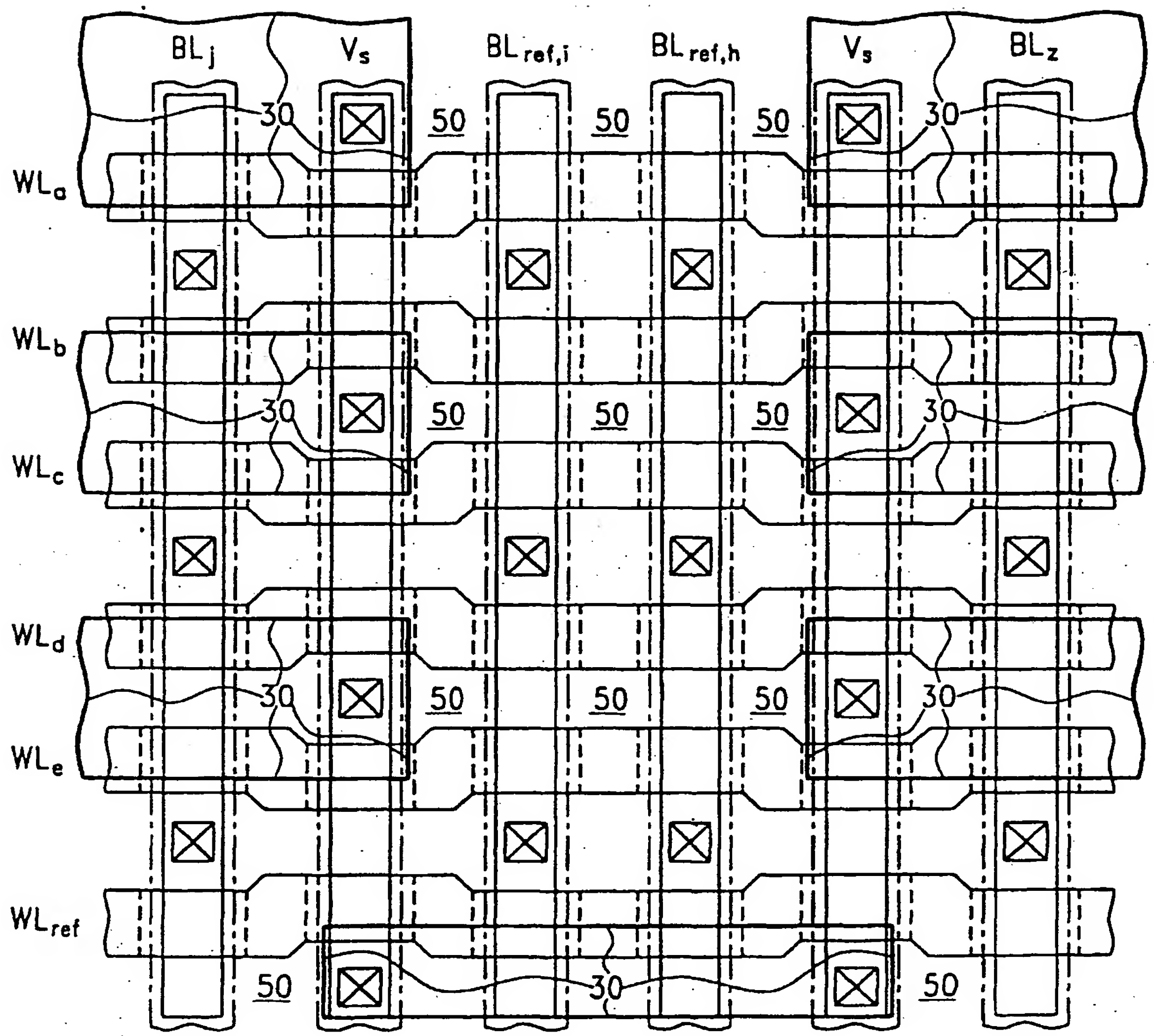


Fig.8

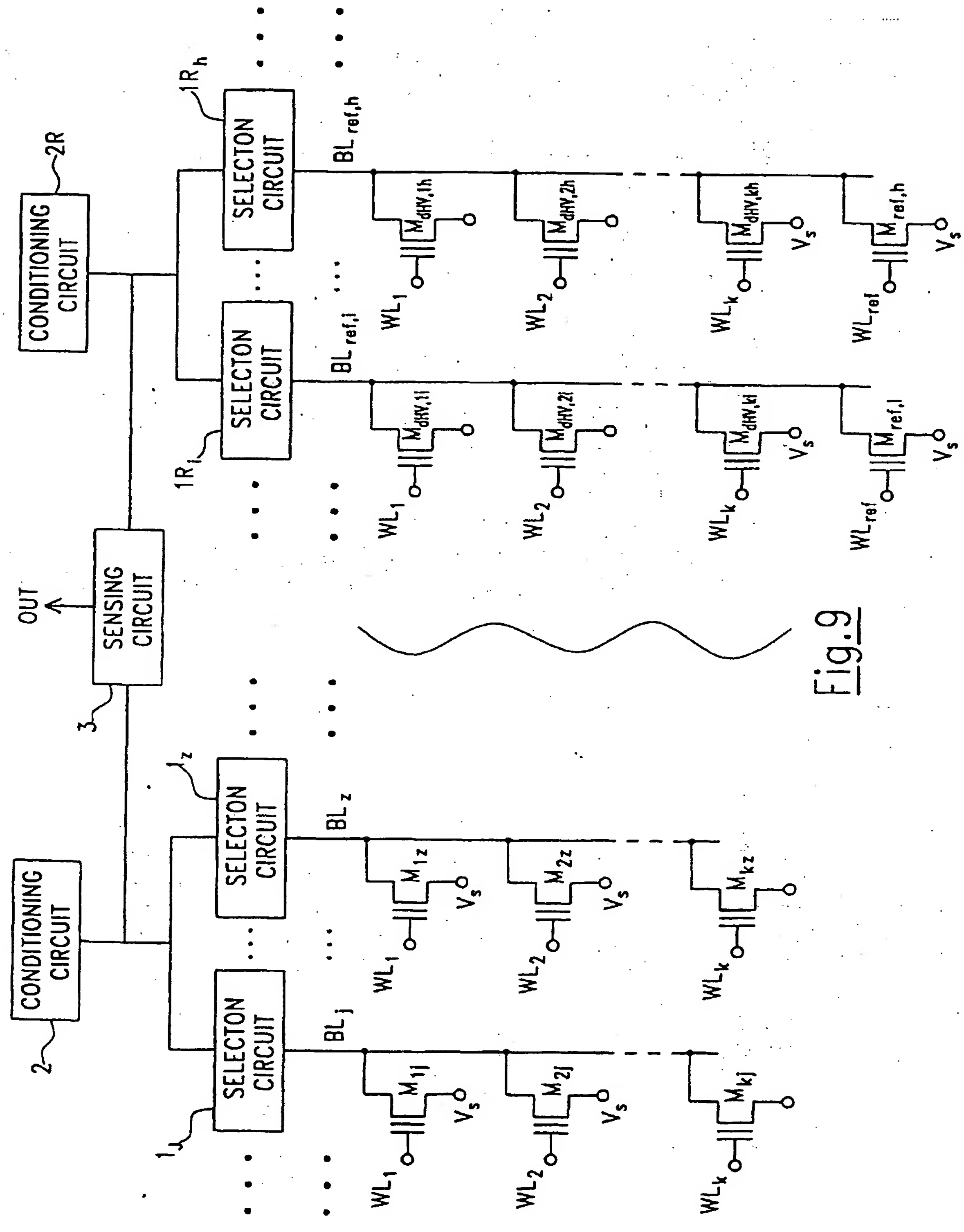
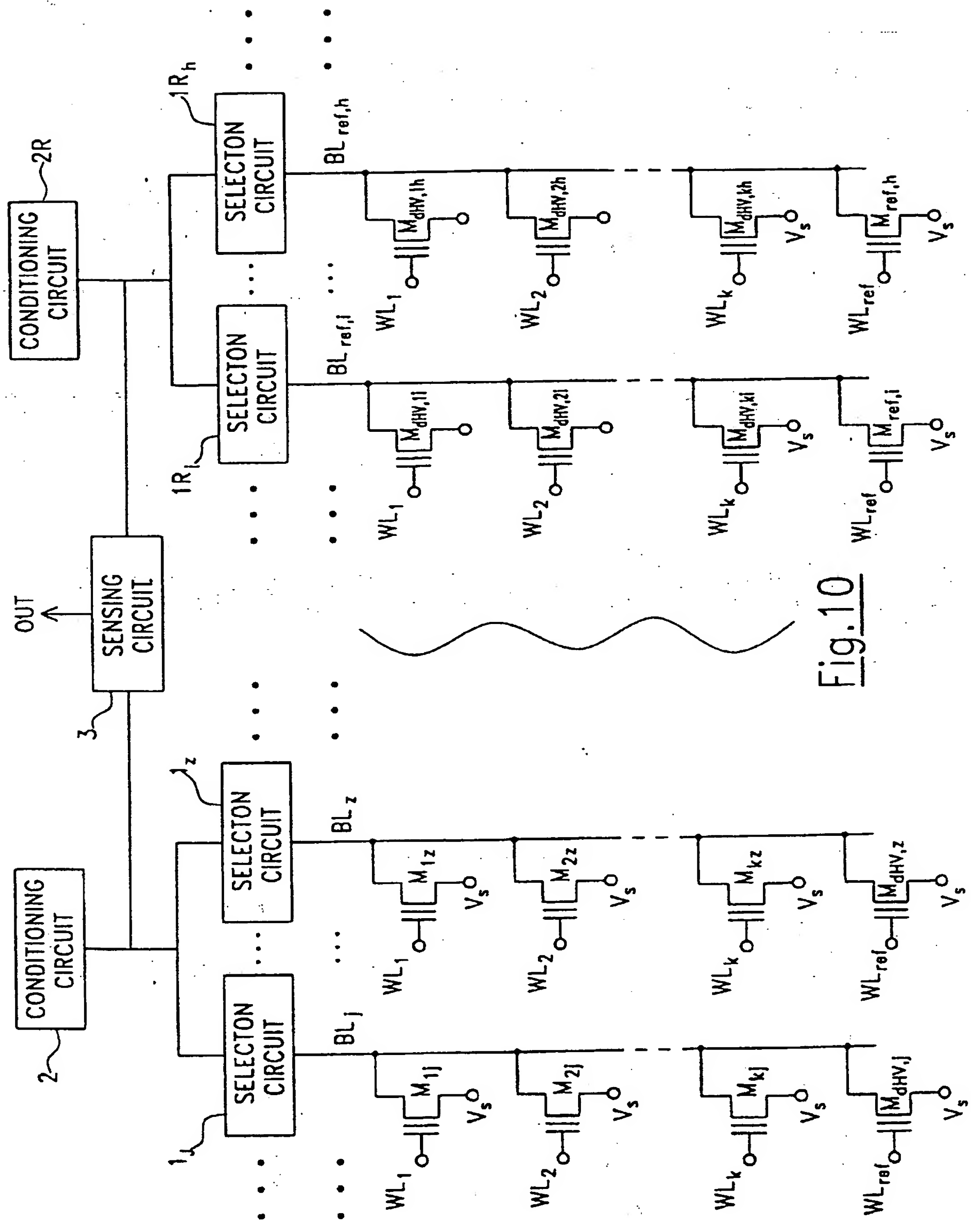

CONDITIONING CIRCUIT
2R
1R_h
SELECTON CIRCUIT
BL_ref,h
M_dHV,1h
M_dHV,2h
M_dHV,kh
M_ref,h
V_s
WL_1
WL_2
WL_k
WL_ref
SELECTON CIRCUIT
1R_i
BL_ref,i
M_dHV,1i
M_dHV,2i
M_dHV,ki
M_ref,i
OUT
SENSING CIRCUIT
3
CONDITIONING CIRCUIT
2
1_z
SELECTON CIRCUIT
BL_z
M_1z
M_2z
M_kz
1_j
SELECTON CIRCUIT
BL_j
M_1j
M_2j
M_kj

Fig.9


CONDITIONING CIRCUIT
2R
CONDITIONING CIRCUIT
2
SENSING CIRCUIT
OUT
3
SELECTON CIRCUIT
1R_h
1R_i
1_z
1_j
BL_ref,h
BL_ref,i
BL_z
BL_j
M_dHV,1h
M_dHV,2h
M_dHV,kh
M_ref,h
M_dHV,1i
M_dHV,2i
M_dHV,ki
M_ref,i
M_1z
M_2z
M_kz
M_dHV,z
M_1j
M_2j
M_kj
M_dHV,j
WL_1
WL_2
WL_k
WL_ref
V_s
Fig.10

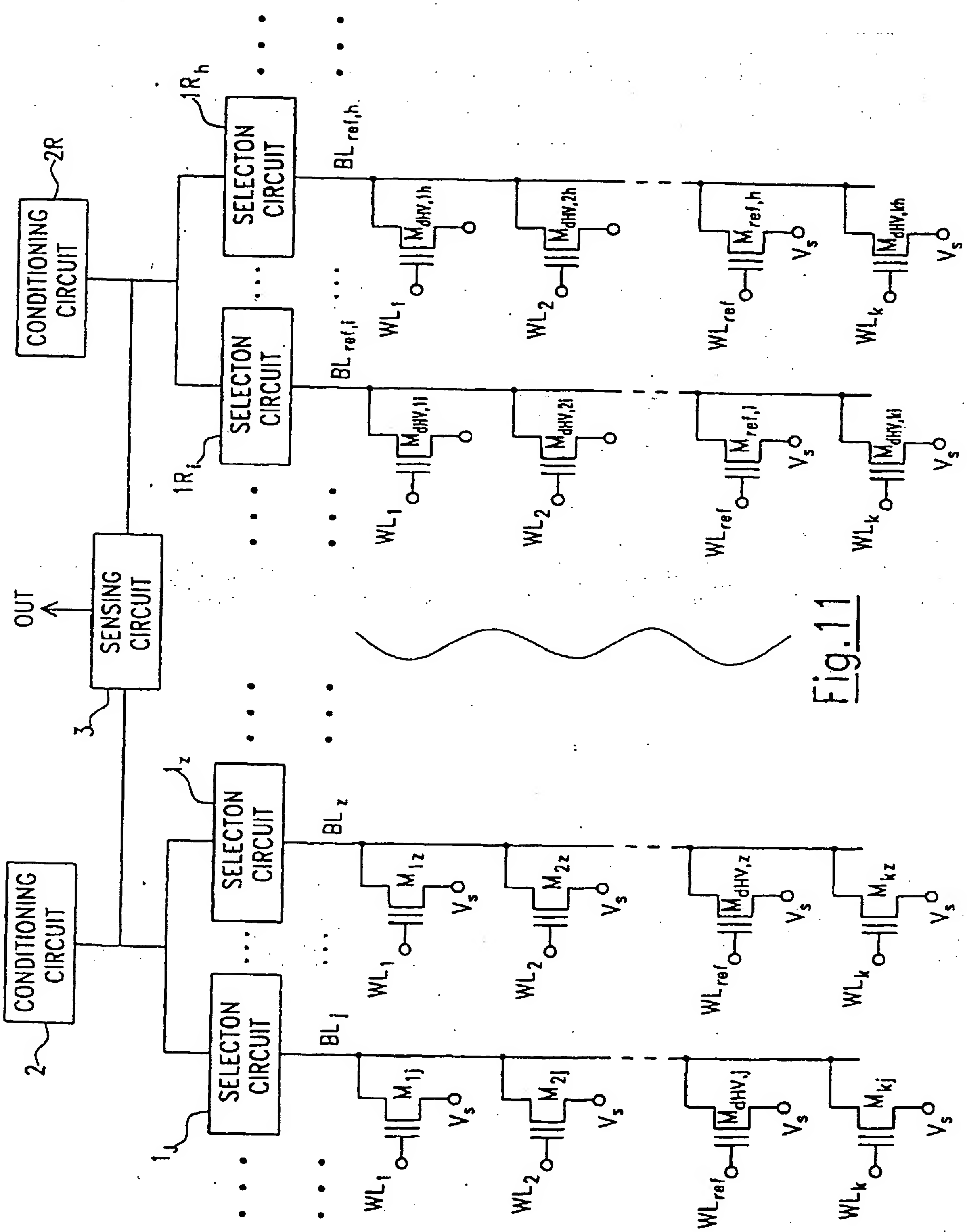

CONDITIONING CIRCUIT
2R
CONDITIONING CIRCUIT
2
SENSING CIRCUIT
3
OUT
SELECTON CIRCUIT
1Rh
1Ri
1z
1j
BLref,h
BLref,i
BLz
BLj
WL1
WL2
WLref
WLk
Vs
MdHV,1h
MdHV,2h
Mref,h
MdHV,kh
MdHV,1i
MdHV,2i
Mref,i
MdHV,ki
M1z
M2z
MdHV,z
Mkz
M1j
M2j
MdHV,j
Mkj

Fig.11

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 98 83 0491

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X Y A	US 5 675 537 A (BILL COLIN STEWART ET AL) 7 October 1997 * figure 3A * * column 5, line 18 - line 21 * * column 7, line 6 - column 8, line 11 *	1,2,9,10 3-8,12 11	G11C11/56
Y	EP 0 337 433 A (TOKYO SHIBAURA ELECTRIC CO) 18 October 1989 * figures 3,4,7,8 * * abstract * * column 4, line 45 - column 6, line 40 * * column 7, line 46 - line 53 *	3-8,12	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
			G11C

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	18 December 1998	Colling, P

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 98 83 0491

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-12-1998

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5675537 A	07-10-1997	WO 9808225 A	26-02-1998
EP 0337433 A	18-10-1989	JP 1262660 A	19-10-1989
		JP 1985712 C	25-10-1995
		JP 7015952 B	22-02-1995
		DE 68917187 D	08-09-1994
		DE 68917187 T	12-01-1995
		US 5105385 A	14-04-1992

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82